Metabolic perturbations prior to hepatocellular carcinoma diagnosis - Findings from a prospective observational cohort study

Stepien, Magdalena; Keski-Rahkonen, Pekka; Kiss, Agneta; Robinot, Nivonirina; Duarte-Salles, Talita; Murphy, Neil; Perlemuter, Gabriel; Viallon, Vivian; Tjønneland, Anne; Rostgaard-Hansen, Agnetha Linn; Dahm, Christina C; Overvad, Kim; Boutron-Ruault, Marie-Christine; Mancini, Francesca Romana; Mahamat-Saleh, Yahya; Aleksandrova, Krasimira; Kaaks, Rudolf; Kühn, Tilman; Trichopoulou, Antonia; Karakatsani, Anna; Panico, Salvatore; Tumino, Rosario; Palli, Domenico; Tagliabue, Giovanna; Naccarati, Alessio; Vermeulen, Roel C H; Bueno-de-Mesquita, H Bas; Weiderpass, Elisabete; Skeie, Guri; Ramón Quirós, J; Ardanaz, Eva; Mokoroa, Olatz; Sala, Núria; Sánchez, Maria-Jose; Huerta, José María; Winkvist, Anna; Harlid, Sophia; Ohlsson, Bodil; Sjöberg, Klas; Schmidt, Julie A; Wareham, Nick; Khaw, Kay-Tee; Ferrari, Pietro; Rothwell, Joseph A; Gunter, Marc; Riboli, Elio; Scalbert, Augustin; Jenab, Mazda

Published in:
International Journal of Cancer

DOI (link to publication from Publisher):
10.1002/ijc.33236

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Metabolic perturbations prior to hepatocellular carcinoma diagnosis – Findings from a prospective observational cohort study

Magdalena Stepien[1]*, Pekka Keski-Rahkonen[1]*, Agneta Kiss[1], Nivonirina Robinot[1], Talita Duarte-Salles[1,2], Neil Murphy[1], Gabriel Perlemuter[3,4,5], Vivian Viallon[1], Anne Tjønneland[6], Agnetha Linn Rostgaard-Hansen[6], Christina C. Dahm[7], Kim Overvad[7,8], Marie-Christine Boutron-Ruault[9,10], Francesca Romana Mancini[9,10], Yahya Mahamat-Saleh[9,10], Krasimira Aleksandrova[11], Rudolf Kaaks[12], Tilman Kühn[12], Antonia Trichopoulou[13,14], Anna Karakatsani[13,15], Salvatore Panico[16], Rosario Tumino[17], Domenico Palli[18], Giovanna Tagliabue[19], Alessio Naccarati[20], Roel C.H. Vermeulen[21], H. Bas Bueno-de-Mesquita[22,23,24,25], Elisabete Weiderpass[1], Guri Skeie[26], J. Ramón Quirós[27], Eva Ardanaz[28,29,30], Olatz Mokoroa[30,31], Núria Sala[32], Maria-Jose Sánchez[30,33], José María Huerta[30,34], Anna Winkvist[35,36], Sophia Harlid[37], Bodil Ohlsson[38], Klas Sjöberg[39], Julie A Schmidt[40], Nick Wareham[41], Kay-Tee Khaw[42], Pietro Ferrari[1], Joseph A. Rothwell[1,10], Marc Gunter[1], Elio Riboli[24], **and** Augustin Scalbert[1≠], Mazda Jenab[1≠]

* These authors contributed equally to this project

≠ These authors contributed equally to this project

1. International Agency for Research on Cancer (IARC-WHO), Lyon, France.
2. Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain.
3. INSERM UMRS U996 – Intestinal Microbiota, Macrophages and Liver Inflammation, DHU Hepatinov, Clamart, France.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.33236

4. Université Paris-Sud, Paris-Saclay, Clamart, France.
5. AP-HP, Hepato-gastroenterology and Nutrition, Antoine-Béclère Hospital, Clamart, France.
6. Diet, Genes and Environment Unit, Danish Cancer Society Research Center, Copenhagen, Denmark
7. Section for Epidemiology, Department of Public Health, Aarhus University, Denmark.
8. Department of Cardiology, Aalborg University Hospital, Denmark.
9. CESP, Faculté de médecine - Université Paris-Sud, Faculté de médecine - UVSQ, INSERM, Université Paris-Saclay, Villejuif, France.
10. Institut Gustave Roussy, Villejuif, France.
11. Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbrücke, Germany.
12. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
13. Hellenic Health Foundation, Athens, Greece.
14. WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Dept. of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Greece.
15. Second Pulmonary Medicine Department, School of Medicine, National and Kapodistrian University of Athens, “ATTIKON” University Hospital, Haidari, Greece.
16. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy.
17. Cancer Registry and Histopathology Department, Provincial Health Authority (ASP) Ragusa, Italy.

18. Cancer Risk Factors and Life-Style Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy.
19. Lombardy Cancer Registry Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy
20. Molecular and Genetic Epidemiology Unit, Italian Institute for Genomic Medicine (IIGM) Torino, Torino, Italy
21. Institute of Risk Assessment Sciences, Utrecht University, The Netherlands.
22. Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands.
23. Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands.
24. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom.
25. Department of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, Pantai Valley, Kuala Lumpur, Malaysia.
26. Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway.
27. Public Health Directorate, Asturias, Spain.
28. Navarra Public Health Institute, Pamplona, Spain.
29. IdiSNA, Navarra Institute for Health Research, Pamplona, Spain.
30. CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain.
31. Public Health Division of Gipuzkoa, Biodonostia Research Institute, San Sebastian, Spain.

32. Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program and Translational Research Laboratory, Catalan Institute of Oncology (IDIBELL), Barcelona, Spain.
33. Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria ibs. Granada. Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain.
34. Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
35. The Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden
36. Department of Public Health and Clinical Medicine, Nutrition Research, Umeå University, Umeå, Sweden
37. Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden.
38. Skåne University Hospital, Department of Internal Medicine, Lund University, Malmö, Sweden.
39. Skåne University Hospital, Department of Gastroenterology and Nutrition, Lund University, Malmö, Sweden.
40. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom.
41. MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom.
42. University of Cambridge, School of Clinical Medicine, Clinical Gerontology Unit, Addenbrooke's Hospital, Cambridge, United Kingdom.

Corresponding authors:

Dr. Mazda Jenab, email: jenabm@iarc.fr Twitter: @Mazda_J and Dr. Augustin Scalbert, email; scalberta@iarc.fr

IARC-WHO, 150 Cours Albert Thomas, Lyon, 69008

List of abbreviations:

ALP, liver-specific alkaline phosphatase

ALT, alanine aminotransferase;

AST, aspartate aminotransferase ;

BMI, body mass index;

DHEA-S, dehydroepiandrosterone sulfate;

EPIC, European Prospective Investigation into Cancer and Nutrition cohort;

GGT, gamma-glutamyltransferase ;

HCC, hepatocellular carcinoma;

HILIC, hydrophilic interaction chromatography;

HPLA, p-hydroxyphenyllactic acid;

lysoPC, lysophosphatidylcholine;

MS, mass spectrometry;

NMR, magnetic resonance spectroscopy;

OR, odds ratio;

PC, phosphatidylcholines;

ROC, Receiver Operation Characteristics curve;

RP, reversed phase chromatography;

SD, standard deviation;

γ-CEHC , γ-carboxyethyl hydroxychroman.

Novelty and Impact:

The aim of this study was to gain insight into metabolic perturbations underlying the development of hepatocellular carcinoma using detailed data from a large, multinational prospective observational cohort. High resolution mass spectrometry-based metabolomics was conducted on blood samples collected pre-diagnostically upon recruitment into the cohort. Cases were identified upon follow-up and compared to matched controls. We controlled for known aetiologies (hepatitis infection, heavy alcohol intake, smoking) and major confounding factors, such as obesity. Alterations were observed in a wide range of metabolites related to exogenous and mutagenic exposures, liver dysfunction and bile acid/phospholipid metabolism, providing insight into early metabolic perturbations and mechanisms leading to this deadly cancer.

Abstract

Hepatocellular carcinoma (HCC) development entails changes in liver metabolism. Current knowledge on metabolic perturbations in HCC is derived mostly from case-control designs, with sparse information from prospective cohorts. Our objective was to apply comprehensive metabolite profiling to detect metabolites whose serum concentrations are associated with HCC development, using biological samples from within the prospective EPIC cohort (>520,000 participants,), where we identified 129 HCC cases matched 1:1 to controls. We conducted high resolution untargeted liquid

chromatography-mass spectrometry based metabolomics on serum samples collected at recruitment prior to cancer diagnosis. Multivariable conditional logistic regression was applied controlling for dietary habits, alcohol consumption, smoking, body size, hepatitis infection and liver dysfunction. Corrections for multiple comparisons were applied. Of 9,206 molecular features detected, 220 discriminated HCC cases from controls. Detailed feature annotation revealed 92 metabolites associated with HCC risk; 14 of which were unambiguously identified using pure reference standards. Positive HCC risk associations were observed for N1-acetylspermidine, isatin, p-hydroxyphenyllactic acid, tyrosine, sphingosine, L,L-cyclo(leucylprolyl), glycochenodeoxycholic acid, glycocholic acid, and 7-methylguanine. Inverse risk associations were observed for retinol, dehydroepiandrosterone sulfate, glycerophosphocholine, γ-carboxyethyl hydroxychroman, and creatine. Discernible differences for these metabolites were observed between cases and controls up to 10 years prior to diagnosis. Our observations highlight the diversity of metabolic perturbations involved in HCC development and replicate previous observations (metabolism of bile acids, amino acids, phospholipids) made in Asian and Scandinavian populations. These findings emphasize the role of metabolic pathways associated with steroid metabolism and immunity and specific dietary and environmental exposures in HCC development.

Keywords: hepatocellular carcinoma; untargeted metabolomics; prospective observational cohort;

Introduction

Primary liver cancer is the second most common cause of death from cancer worldwide (1). Established risk factors for hepatocellular carcinoma (HCC), the major histology of primary liver cancers, are chronic hepatitis infection, aflatoxin exposure, smoking and alcohol abuse (2), but obesity, diabetes and unhealthy dietary and lifestyle habits are also becoming increasingly recognized as important HCC risk factors, particularly in regions where hepatitis infection and aflatoxin exposures are less predominant (3). HCC are often diagnosed at late stages and have limited treatment options, which is worrisome owing to the growing incidence of this highly fatal disease in many populations (4). It has been suggested that high obesity and diabetes rates in some populations are major contributors to the observed incidence rate increases (5). Most HCC are considered to develop within a background of inflammation, liver damage and cirrhosis. However, a sizeable proportion is thought to develop in the absence of underlying cirrhosis, hence escaping traditional clinical surveillance particularly in populations with lower prevalence of hepatitis infection and alcohol abuse, and higher prevalence of metabolic syndrome and non-alcoholic fatty liver disease (NAFLD) which are largely obesity-related (6;7). Obesity may also impair the detection of cirrhosis or HCC by reducing the sensitivity of abdominal ultrasound, a primary tool for HCC surveillance in high risk populations (8). Thus, effective HCC control will need to rely on strategies for both primary prevention and early detection, necessitating additional research into HCC etiology.

Decreased liver functionality is considered an early event in liver cancer development and given the central metabolic role of the liver various metabolic perturbations are

very likely to be observed in blood. In addition, circulating biomarkers indicative of various lifestyle or environmental exposures that may affect HCC risk are also likely observable (9). Such metabolic signatures can be identified via various metabolomic techniques, such as those based on high resolution liquid chromatography mass spectrometry (LC-MS), which may be applied to blood samples to observe a broad spectrum of low-molecular-weight compounds which may be reflective of various exogenous exposures and associated with normal endogenous processes or perturbed metabolic functionality. In fact, several animal and human studies have already shown that metabolomics can provide novel insights into pathological processes during development of various liver diseases (10;11), and provide potentially novel diagnostic biomarkers of HCC for screening in high risk populations (12-14). Most of the studies that have applied metabolic profiling in HCC have been either based on case-control designs, or conducted on high risk patient groups (e.g. viral hepatitis, cirrhosis or other chronic liver diseases), or in populations where more traditional HCC risk factors predominate (15). However, comparatively very little information is available from prospective, observational cohorts about possible metabolic alterations related to HCC development, particularly from European or Western populations (16-18). Information derived from prospective observational cohorts is important because data and biological samples have been collected from healthy participants before diagnosis, thus reducing the biases of recall and reverse causality and allowing considerable insight into the complex processes of cancer development. For example, within the European Prospective Investigation on Cancer

and Nutrition (EPIC) cohort, a number of targeted metabolomic studies (i.e. the measurement of defined groups of characterized and annotated metabolites; about 150 metabolites measured) have been conducted to assess metabolite patterns associated with risk of several cancers such as the breast (19) and prostate (20;21), as well as with various lifestyle factors, such as body mass index (22) and select dietary components (23). They have revealed important insights on development processes and exogenous exposures associated with these cancers. Similar metabolomics techniques have also been applied in other prospective studies to explore cancer development at various anatomical sites, including HCC (18;24-26). We have also previously conducted two other metabolomics studies on HCC risk factors in the EPIC cohort using nuclear magnetic resonance spectroscopy (NMR) (17) and a targeted kit-based LC-MS assay (16). We observed alterations in amino acid, lipid and carbohydrate metabolism associated with HCC development, but our findings provided little new insight into HCC etiology or specific environmental exposures potentially linked to HCC development, due in large part to the low sensitivity of NMR (17) and the limited number of metabolites measured with the kit-based assay (16).

In the present study, our objective was to delve more deeply into an exploration of metabolic perturbations in HCC development through application of untargeted metabolomics (i.e. the comprehensive analysis of all measurable analytes, but requiring intensive efforts towards metabolite annotation) using a highly sensitive LC-

MS technique able to detect thousands of metabolites in typical blood samples (27) using a case-control design nested within the prospective EPIC cohort.

Materials and methods

Study design

The rationale and study design of the large multi-center prospective, observational EPIC cohort have been previously described (28). Briefly, between 1991 and 2000 more than 520,000 apparently healthy men and women aged 20-85 years were recruited in 23 centers throughout 10 countries (Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom). At recruitment, standardized dietary, lifestyle and socio-demographic questionnaires, blood samples and anthropometric measurements were collected from most participants (29). Blood samples are stored at the International Agency for Research on Cancer (IARC-WHO, Lyon, France) in -196°C liquid nitrogen for all countries except Denmark (-150°C, nitrogen vapour) and Sweden (-80°C, freezers), where they are stored locally.

Nested Case-Control Study

From 477,206 eligible participants, we included 129 HCC cases (diagnosed post-recruitment into the cohort and identified up to December 2010) with available baseline (i.e. pre-diagnostic) blood samples. The cases were followed-up for a median of 6.2 years / mean of 5.9 years from baseline recruitment until HCC diagnosis. For each case, we selected one control (n=129) by incidence density

sampling from all eligible cohort participants alive and matched by age at blood collection (±1 year), sex, study center, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6,and >6 hours); and additionally among women by menopausal status (pre-, peri-, and postmenopausal), and hormone replacement therapy use at time of blood collection (yes/no). Incidence density sampling for control selection is a common method of choice for unbiased results in case-control studies nested within a prospective cohort (30). The method involves matching each case to a sample of those who are at risk from within the cohort population at the time of case occurrence.

HCC was defined as C22.0 according to the 10th revision of the International Statistical Classification of Diseases, Injury and Causes of Death (ICD10), with morphology codes “8170/3” or “8180/3” according to the 2nd edition of the International Classification of Diseases for Oncology (ICD-O-2). For each case identified, the histology and diagnostic methods were reviewed by a trained pathologist to exclude metastatic cases or other types of primary liver cancers. Details on participant exclusion criteria and cancer incidence determination are described in the **Supplementary Materials and Methods**.

Untargeted Metabolomics

Detailed methods for the metabolomics analyses (i.e. sample preparation and analysis. data preprocessing, and feature identification) are provided in the **Supplementary Materials and Methods**. Briefly, samples were analysed with a

UHPLC-QTOF-MS system (Agilent Technologies, Santa Clara, CA, USA) using four different analytical configurations with reversed phase (RP) or hydrophilic interaction chromatography (HILIC) columns and positive or negative MS ionization modes (i.e. RP +/-, HILIC +/). Peak areas were used as a measurement of feature intensity. For identification, mass to charge rations (*m/z*) were searched against the Human Metabolome Database (31) and METLIN (32), using ions [M+H]+, [M+Na]+, [M-H]-, [M+FA-H]-, with 8 ppm molecular weight tolerance. Where pure chemical standards were commercially available, identification was confirmed by reanalysis of representative samples and pure chemical standards comparing retention times and MS/MS spectra. When standards were not available, MS/MS spectra were acquired when possible and compared against those in mzCloud (www.mzcloud.org) or METLIN. Level of identification was determined as proposed by Sumner et al (33) in line with recommendations of the Metabolomics Standards Initiative which ranks metabolites into 4 distinct categories: unambiguous identification using pure standards (Level 1), identified with a high level of confidence based on chemical features and characteristics (Level 2), identified to a known chemical class (Level 3) and unknown / unidentifiable compounds (Level 4). For the purposes of this analysis, Levels 1-3 are considered identified metabolites, but with varying levels of certainty (i.e. unambiguous, highly likely and chemical class only).

Additional Laboratory Measures

In a large subset of subjects, a score of liver function (indicator of underlying liver damage) was computed using additional and already available biomarker measures (34;35) (details in Table 1 footnotes).

Dataset preparation and statistical analyses

A separate analysis was conducted for each dataset from the four analytical configurations (i.e. RP +/-, HILIC +/-). In each dataset, features missing from more than25% of all samples were excluded to avoid extensive imputation of the data before the paired statistical analysis (see **Figure 1** for details). In order to retain a maximum number of complete case-control sets in the statistical analyses, missing values for any feature (features not detected in a given subject) for any feature were replaced with the minimum intensity of that feature in the dataset (**Figure 1**). To assess differences between cases and controls, feature intensities were log2-transformed (to improve data normality) and z-standardized (to better enable comparisons across a wide intensity range), and subsequently entered into conditional logistic regression models from which odds ratios (OR) and 95% confidence intervals (95% CI) were computed. Two main statistical models were applied, **(a)** a **crude model**, conditioned on the matching criteria only and **(b)** a detailed **multivariable model** with additional adjustments for continuous variables body mass index (BMI, kg/m^2), waist circumference (cm), recreational and household physical activity (Met-hours/week), alcohol intake at recruitment (g/d), and categories of lifetime alcohol intake pattern, smoking status and highest level of education attainment (for categories see **Table 1**). The Benjamini-Hochberg correction for

multiple testing was applied using the multi-test procedure in SAS and a q-value of ≤0.05 was considered as statistically significant. Additionally, fold change between the median intensity for the cases *vs.* the controls was used to rank the features by their absolute intensity difference. Thresholds for the selection of the most discriminating features for annotation were based on absolute median fold change of ≥1.20. Additional adjustments for hepatitis infection status (to correct for this established risk factor), self-reported type-2 diabetes at baseline (to correct for potential influence of diabetes-related metabolic dysfunction) and a composite score of liver function (to correct for the extent of liver dysfunctionality and capacity) were applied in supplementary analyses for all identified features.

Sensitivity analyses were conducted excluding first 2 and 4 years of follow-up (n=22 and n=43 cases excluded, respectively) to assess potential reverse causation. For these analyses, a p-value ≤0.05 was considered statistically significant.

Pearson correlation coefficients were used to assess the correlations between metabolites that were annotated (i.e. those at Levels of identification 1-3(33), but not unknown metabolites). For these same annotated metabolites, we conducted principal component analyses in order to illustrate the separation of profiles from baseline over the timeline of the follow-up period (i.e. from baseline recruitment into the cohort to the date of diagnosis) between identified features of cases and controls. In addition, we then constructed a Receiver Operating Characteristics (ROC) curve based on stepwise forward selection of metabolites from the panel of metabolites that were annotated to Level 1 (i.e. the panel of metabolites that were significantly

different between cases and controls and unambiguously identified using a pure standard) and those at Levels 1-3 (33). The final areas under the ROC curve for the identified discriminant features were obtained using leave-one-out cross validation.

All statistical tests were two-sided. Analyses were conducted using SAS version 9.3 (SAS Institute, Cary, NC), R version 3.4.3 (Principal Component Analyses) or MetaboAnalyst version 4.0 (Heatmap).

Results

Characteristics of the HCC cases and their matched controls are presented in **Table 1**. Cases were primarily men, former drinkers and current smokers, and had higher waist circumference, higher prevalence of hepatitis B/C infection, and higher degree of liver dysfunction than matched controls. The average length of follow-up was 8.5 years for cases and controls combined and 5.9 years for cases alone, with a maximum follow-up length of 15 years from baseline.

From the combined total of 9,206 molecular features provided by the four analytical configurations of the LC-MS, 5,229 (i.e. 2551 (RP+), 1178 (RP-), 736 (HILIC+) and 764 (HILIC-)) were present in at least 75% of all samples and were exported for statistical analyses (**Figure 1**). Initially, 333 (RP+), 20 (RP-), 68 (HILIC+) and 14 (HILIC-) features were found to be statistically significantly associated with HCC risk in multivariable models (**Supplementary Tables 1A, 1B, 1C and 1D, respectively**). Excluding the features with a median fold change less than 1.20 resulted in a total of 220 features from the 4 analytical configurations combined. From these 220 features,

114 individual compounds (i.e. confirmed molecules that consisted of one or more features) were observed in the four datasets (**Figure 1**) and are visualized in volcano plots (**Supplementary Figure 1**).

Of the 114 individual compounds, 22 were also detected by at least one of the other three profiling configurations, leaving a total of 92 unique annotated compounds. Each profiling configuration identified at least 5 unique compounds, highlighting the advantages of applying all four orthogonal analytical configurations for more comprehensive metabolite coverage. Identification was attempted for each of these 92 unique compounds, ranking them according to varying levels of confidence based on the recommendations of the Metabolomics Standards Initiative (33). Fourteen metabolites were unambiguously identified using pure reference standards (Level1, **Table 2**), another 23 compounds were identified with a high level of confidence based on chemical features and characteristics (Level 2, **Table 3**) and 9 compounds were identified to a known chemical class (Level 3, **Table 4**), summing to a total of 46 identified metabolites. Pearson correlation coefficients between these 46 identified metabolites are shown in **Supplementary Figure 2**. The remaining 46 metabolites could not be identified and are listed as unknown (Level 4, **Table 4**). Intensity means, standard deviation and medians for these 92 compounds are shown in **Supplementary Table 2A**.

The multivariable adjusted HCC risk associations for the 14 metabolites identified at Level 1 are shown in **Table 2**. Of these 14 metabolites, 5 were inversely associated with HCC risk (q-value ≤0.05): retinol (OR=0.27, 95%CI: 0.16-0.48),

dehydroepiandrosterone sulfate (DHEA-S; OR=0.35, 95%CI: 0.22-0.57), glycerophosphocholine (OR=0.44, 95%CI: 0.28-0.71), γ-carboxyethyl hydroxychroman (γ-CEHC; OR=0.56, 95%CI: 0.39-0.81), creatine (OR=0.56, 95%CI: 0.37-0.83). The remaining 9 metabolites were positively associated with HCC risk: N1-acetylspermidine (OR=2.16, 95%CI: 1.38-3.37), isatin (OR=2.56, 95%CI: 1.53-4.29), p-hydroxyphenyllactic acid (HPLA; OR=2.63, 95%CI: 1.62-4.28), tyrosine (OR=2.77, 95%CI: 1.58-4.83), sphingosine (OR=2.79, 95%CI: 1.66-4.71), L,L-cyclo(leucylprolyl) (OR=3.25, 95%CI: 1.91-5.53), glycochenodeoxycholic acid (OR=3.31, 95%CI: 1.99-5.51), glycocholic acid (OR=4.07, 95%CI: 2.32-7.14), and 7-methylguanine (OR=6.78, 95%CI: 3.24-14.18).

We additionally conducted ROC discriminant analyses from the panel of the 14 Level 1 identified metabolites. The analyses showed that the discrimination between cases and controls was largely driven by retinol, DHEA-s, LL-cycloleucylpropyl and 7-methylguanine. Additional ROC analysis using leave-one-out cross validation for these 4 independent metabolites indicated a 84.6% discriminatory accuracy, compared to a 85.0% discriminatory accuracy when all 14 Level 1 identified metabolites were modelled. This method of validation was chosen to avoid likely statistical power issues that would arise from splitting the main dataset into discovery and validation sub-sets, each of which would include a smaller number of cases. Conversely, we applied the leave-one-out cross validation approach to the identified metabolites rather than at the stage of feature selection, as would be the case in a

true validation setting with training and validation sub-sets. Thus, the AUC estimate is likely to be biased.

Multivariable-adjusted HCC risk associations for the twenty-three Level 2 metabolites (largely phosphatidylcholines (PC), lysophosphatidylcholines (lysoPC) of various chain lengths, diacylglycerols, two bilirubin metabolites and benzoylcarnitine) are shown in **Table 3**. Multivariable-adjusted HCC risk associations for Level 3 (some glycerophosphocholines and C19 steroid sulfates) and Level 4 compounds are shown in **Table 4**.

Results for the crude models conditioned on the matching criteria only are shown in **Supplementary Table 2B**. Supplementary analyses with additional adjustments for hepatitis B and/or C infection status, self-reported diabetes status at baseline (**Supplementary Table 2C**), and a score of liver functionality within the multivariable analysis model did not materially alter the findings **(Supplementary Table 2D).** In sensitivity analyses, the observed associations, particularly for Level 1 and Level 2 compounds were unaltered after exclusion of case-control pairs where the case participant was diagnosed within either the first 2 or 4 years of follow-up **(Supplementary Table 2E**).

We conducted two principal component analyses, one based on the 46 metabolites identified to Levels 1-3 as well as a second one restricted to the 14 metabolites identified to Level 1 (**Figure 2**). These analyses show distinct differences between metabolic profiles of HCC cases compared to control participants, up to 10 years

prior to diagnosis. Additional sensitivity analyses excluding case-control pairs where the case was diagnosed within the first 4 years of enrolment into the cohort did not alter the clear distinction in metabolite profiles between HCC cases versus controls (**Figure 2**).

Detailed information on metabolite identification with chromatograms and spectra is provided in **Supplementary Materials (Identification of Metabolites)**.

Discussion

In this case-control study nested within a large, multinational observational prospective cohort, we applied a powerful MS-based untargeted metabolomics approach to explore metabolic perturbations underlying HCC development. The cases in our observational cohort were enrolled at the baseline period (i.e. data and blood samples collected upon recruitment) when the participants where under apparent health. Later, at various time points post-recruitment, some of the cohort participants were diagnosed with HCC. Thus, the cases in our study originate from different time points after baseline recruitment. In the sub-group of cases who were diagnosed closer to baseline, it is likely that the processes of HCC were already underway even though undiagnosed, possibly within a background of other liver pathologies. However, in the sub-group of subjects where the HCC was diagnosed later on during the cohort follow-up, the baseline blood samples are likely to have been collected in the absence of HCC or at its earlier stages. Due to the liver's central metabolic roles, it is thought that metabolic disturbances are early events in

the development of chronic liver diseases and HCC (36). This premise underscores the rationale behind our study, conducted within the setting of an observational prospective cohort. We were able to determine 92 distinct metabolites whose relative concentrations were different between HCC cases and their matched controls in pre-diagnostic blood samples. Of these 92 compounds, we were able to identify 46 of which 14 were unambiguous (Level 1 (33)) and an additional 23 and 9 with high degrees of confidence (Levels 2 and 3, respectively (33)). We show, using principle component analyses, that the differences between HCC cases and controls are apparent as far back as 10 years prior to diagnosis, even with exclusion of cases diagnosed within the first 2 or 4 years of follow-up. We observed perturbations in general classes of metabolites, such as amino acids and bile acids, but also in xenobiotics as indicators of lifestyle exposures, as well as some compounds with purported roles in immune function, hormone metabolism, gut microbiome activity and liver fat content - underscoring the complexity of metabolic disturbances in HCC development. The metabolites identified may be involved directly and/or be markers of various exposures associated with cancer risk. Moreover, we accounted for established etiologies of HCC such as hepatitis infection, high alcohol consumption and smoking in our statistical analysis models. Our observations were mostly unchanged with these adjustments, suggesting that metabolic perturbations in HCC may be largely similar, irrespective of the main underlying etiology of the tumor.

Of the 46 metabolites that we could identify in this study, 14 were confirmed using authentic chemical standards. Several of these appear to be related to dietary and

lifestyle habits. Specifically, we observed inverse HCC risk associations for retinol (biologically active form of vitamin A) and γ-CEHC (a product of liver metabolism of γ-tocopherol) (37;38). Retinol has a plausible role in liver carcinogenesis (e.g. modulation of immune function, cell growth (39)). Its potential association with liver cancer has been previously assessed in two prospective studies, a Finnish cohort of male smokers (40) as well as a cohort of Chinese men (41) both of whose findings are in line with our own observations. γ-CEHC shows some antioxidant and anti-inflammatory properties, similar to γ-tocopherol (38;42). It has been purported as a treatment of non-alcoholic steatohepatitis, a precursor of liver cirrhosis and risk factor for HCC development (43), but little other data is available on any specific HCC protective roles for this compound.

We also observed inverse HCC risk associations for glycerophosphocholine, several lysoPCs, creatine and DHEA-S, a steroid hormone. Interestingly, decreased glycerophosphocholine level has been observed to be predictive of higher circulating vitamin D concentrations (44), which would be in line with our earlier observation of a strong inverse HCC risk association with higher circulating vitamin D in these same subjects (45). Inverse HCC risk associations with higher circulating lysoPCs are consistent with other reports (10;15;46). The observed association with creatine may reflect decreased liver functionality and lower creatine synthesis in HCC development, although it has also been ascribed both antioxidant and oxidative properties (47). Our observation of an inverse association with DHEA-S is intriguing because androgen receptor activity, with which DHEA-S interacts, has been

implicated in HCC development (48;49) and the promotion of HCC by androgens has been put forward as one explanation for its higher incidence in men (50). On the other hand, liver cirrhosis has been linked to hormonal imbalances between estrogens and androgens resulting in a higher relative concentration of estrogens (50). Some animal data even suggest that DHEA-S may protect against development of liver lesions (51). Thus, our observations merit more detailed assessment of hormonal factors and circulating concentrations.

In an earlier study based on NMR spectroscopy within the same subjects, we found a positive HCC risk association for the amino acid tyrosine (16). Similar observations have been made in a Korean prospective cohort (25) and the Alpha-Tocopherol, Beta-Carotene Cancer prevention cohort (ATBC) composed of Finnish male smokers (18). Our observations in the present study were similar for tyrosine along with HPLA, a tyrosine metabolite. Tyrosine is found in several foods (e.g. cheeses, which incidentally have also been associated with increased HCC risk in our data (52)) and is produced endogenously from phenylalanine. Tyrosine levels are known to be altered in liver disease (53) while HPLA has demonstrated carcinogenic activity after long term sub-cutaneous injection in mice (54), and its urinary levels have been observed to be elevated in breast cancer patients (55).

Positive HCC risk associations were also observed for isatin, L,L-cyclo(leucylprolyl), N1-acetylspermidine and sphingosine – although, very little is known about any physiological roles for these compounds in HCC development. Isatin is a biologically active endogenous metabolite with antioxidant and antiviral effects (56) - properties

that may be considered as cancer protective rather than explanatory of our observed positive HCC risk association. However, isatin can also be derived from gut microbial metabolism (56), and we can speculate that its higher circulating concentrations in HCC cases may be due to leakage from the gut across a dysfunctional colonic barrier, something which we have previously observed in the same HCC cases (57). For its part, L,L-cyclo(leucyl-prolyl) has been associated with increased liver fat content in a German general population sample (58), possibly suggesting a link with fatty liver disease in some of our cases. There is sparse data on the possible roles of N1-acetylspermidine (a polyamine) and sphingosine (an aminodiol which can form ceramides, parent structures to sphingolipids) in HCC. The former may be affected by liver functionality (59) and its serum levels have been shown to be higher in liver cancer patients (60) whereas sphingosine has been observed to be elevated in chronic liver diseases, such as non-alcoholic fatty liver disease and chronic hepatitis C infection (61). Thus, alterations in circulating levels of these metabolites may be indicative of liver dysfunctionality and possibly early HCC development.

Another interesting observation from our study is the positive HCC risk association of glycochenodeoxycholic acid and glycocholic acid – both of which are glycine conjugates of primary bile acids formed in the liver (62). Their circulating concentrations have been shown to be increased in various liver diseases, including HCC (10;15;25;46;63;64). In general, pro-inflammatory and carcinogenic properties have been ascribed to bile acids and as such they have a plausible role in HCC development (65). Perturbations in serum bile acid metabolism have been previously

observed in other settings, such as in largely hepatitis positive Chinese populations (66;67), and specifically for glycochenodeoxycholic acid and glycocholic acid in the ATBC cohort (18). The liver is central to bile acid metabolism, and hence perturbations in bile acid profiles may be amongst the earliest indicators of HCC development. A more detailed analysis of potential alterations in the profiles of various bile acids in the different phases of this disease would be of great interest.

In our observations, the metabolite most strongly positively associated with HCC risk is 7-methylguanine, an indicator of exposure to methylating agents. It has previously been observed to be higher in the urine of smokers (68) and those with unhealthy lifestyle habits (69) – exposures which have also been associated with increased HCC risk in our cohort (34;70). Higher levels of this compound have also been associated with an increased risk of total mortality in a cohort of male smokers (71). It may thus be a metabolite related to smoking exposure, and so further study of its potential role in HCC development is warranted.

Taken together, our findings relate to dietary and lifestyle exposures that may be potentially HCC promoting, as well as to liver dysfunctionality which is central to the development of HCC and other liver diseases.

A major limitation of our study nested within a prospective cohort is the lack of information on the existence and severity of any other liver diseases leading up to HCC development. For example, information on existing liver cirrhosis would have been helpful in further characterizing our HCC cases between cirrhotic and non-

cirrhotic pathways of HCC development. A related critique of our study design is the lack of a second control group composed of subjects with liver diseases. We do not have any access to relevant clinical information on liver diseases amongst our >520,000 cohort participants. However, such a control group would have allowed us to better understand transitions from existing liver pathologies towards early HCC. Although this is a reasonable assertion for studies designed to assess clinical surveillance for HCC in higher risk populations, it is less relevant to prospective cohorts geared towards exploring cancer etiology in the general population. Nevertheless, we have addressed these concerns by making multivariable statistical adjustments for main HCC risk factors in our study population. These adjustments did not meaningfully alter our findings, suggesting that different HCC etiologies – whether related mainly to chronic hepatitis infection, alcohol abuse, smoking or obesity – may have a large degree of overlap in terms of their metabolic consequences on the liver and hence transitions towards development of HCC. Patient cohorts comparing HCC cases to control subjects with liver disease provide vital insight towards risk stratification for HCC screening and identification of diagnostic biomarkers, they have to be distinguished from findings such as ours which are based on large-scale prospective cohort studies and which bring understanding of potential risk factors and metabolic perturbations in HCC development. Another limitation is our lack of information on any tumour staging criteria at diagnosis or treatments post-diagnosis. We did not consider survival and we cannot discount some degree of confounding by stage at diagnosis – but it must

be noted that biological samples in our cohort were collected at recruitment, pre-diagnosis. In randomized clinical trials, allocation of exposure and prognostic factors would be random, but collection of data and biological samples in the cases is not likely to be pre-diagnostic. Studies with these different designs each provide crucial insight into the development of this lethal cancer, and should all be part of the evidence base for establishment of guidelines towards HCC prevention, as well as discovery of biomarkers for early diagnosis. We consider the fact that our HCC cases were derived from within an observational cohort with pre-diagnostically obtained biological samples and detailed confounder data as a major advantage that minimises recall and reverse causality biases adding another degree of robustness to our observations. At the same time, we acknowledge that our study design does not allow insight into transitions from existing liver pathologies towards HCC.

Another important design advantage of this work is that we applied an agnostic metabolomics approach using high-resolution mass spectrometry with four complementary analytical configurations (72) enabling us to maximize the number of metabolites measured for a more complete assessment of metabolic profile changes between the HCC cases and their matched controls. We identified many metabolites with very high confidence, but we also observed a number which we could not identify despite our best efforts. We believe that the high number of metabolites observed to be associated with HCC risk, both identified and unknown, highlights the depth of metabolic perturbation in this disease. The magnitude of some of the risk associations for the unidentified metabolites, whether inverse or positive, shows that

we still have much to learn about the processes of HCC development. The unidentified metabolites provide considerable potential for discovery of additional novel exposure, diagnostic and prognostic biomarkers in other studies. Our findings on specific identified metabolites and metabolic pathways involved in HCC development may be followed up with experimental studies to more carefully query their functionality and mechanisms of action. Additionally, it would be of great interest to determine whether any of our observed metabolites may serve as early diagnostic markers.

In summary, we show statistically significant associations between 46 identified metabolites, which could be either directly involved in HCC development or be the consequence of liver dysfunction caused by tumourigenesis in the liver. Our observations, based on pre-diagnostically collected blood samples, contribute towards a more in-depth understanding of HCC risk factors and underlying mechanisms of HCC development. They contribute to the evidence base that may be used towards public health guidelines for HCC prevention, but they should also be replicated in other prospective cohorts from different world regions with emphasis on comparing metabolic changes over time from the earliest phases of HCC development.

Declarations:

Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Ethics statement

All cohort members provided written informed consent. Approval for this study was obtained from the relevant ethical review boards of the participating institutions and from the IARC Ethics Committee.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the French National Cancer Institute (L'Institut National du Cancer; INCA) (grant number 2014-1-RT-02-CIRC-1; PI: M. Jenab). The coordination of EPIC is financially supported by the European Commission (DG-SANCO); and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer; Institut

Gustave Roussy; Mutuelle Générale de l'Education Nationale; and Institut National de la Santé et de la Recherche Médicale (INSERM) (France); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum (DKFZ); and Federal Ministry of Education and Research (Germany); Hellenic Health Foundation (Greece); Italian Association for Research on Cancer (AIRC); National Research Council; and AIRE-ONLUS Ragusa, AVIS Ragusa, Sicilian Government (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS); Netherlands Cancer Registry (NKR); LK Research Funds; Dutch Prevention Funds; Dutch ZON (Zorg Onderzoek Nederland); World Cancer Research Fund (WCRF); and Statistics Netherlands (the Netherlands); and Nordic Center of Excellence Programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS); Regional Governments of Andalucía, Asturias, Basque Country, Murcia (No. 6236) and Navarra; and ISCIII RETIC (RD06/0020) and the Catalan Institute of Oncology (Spain); Swedish Cancer Society; Swedish Scientific Council; and Regional Government of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 for EPIC-Norfolk and C570/A16491 for EPIC-Oxford) and the Medical Research Council (1000143 for EPIC-Norfolk and MR/M012190/1 for EPIC-Oxford) (UK).

The funding sources had no influence on the design of the study; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the paper for publication.

Authors' contributions: The authors' responsibilities were as follows. MJ: conceptualized, designed, obtained funding for and implemented the present

research; NR: conducted LC-MS analyses; PK-R: performed metabolomics data processing; PK-R and AK: performed metabolites identification; MS and PK-R: performed the statistical analysis; VV and PF: provided input and advice on the statistical analysis strategy; MS, MJ, AS, PK-R: contributed jointly to data interpretation and writing of the manuscript. MG, TDS, GP and NM: provided input and critical comment on data interpretation and manuscript writing. Contributing authors from each individual collaborating center provided the original data and biological samples, information on the respective populations, advice on study design/analysis, and interpretation of the results. All authors provided an approval of the final version of the manuscript for publication. The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

Acknowledgements

Not applicable

Data Accessibility

For information on how to submit an application for gaining access to EPIC data and/or bio-specimens, please follow the instructions at http://epic.iarc.fr/access/index.php.

References

(1) Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424.

(2) Kew MC. Hepatocellular carcinoma: epidemiology and risk factors. J Hepatocell Carcinoma 2014;1:115-25.

(3) Blachier M, Leleu H, Peck-Radosavljevic M, Valla DC, Roudot-Thoraval F. The burden of liver disease in Europe: a review of available epidemiological data. J Hepatol 2013 Mar;58(3):593-608.

(4) McGlynn KA, Petrick JL, London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. Clin Liver Dis 2015 May;19(2):223-38.

(5) Kulik L, El-Serag HB. Epidemiology and Management of Hepatocellular Carcinoma. Gastroenterology 2019 Jan;156(2):477-91.

(6) Trevisani F, Frigerio M, Santi V, Grignaschi A, Bernardi M. Hepatocellular carcinoma in non-cirrhotic liver: a reappraisal. Dig Liver Dis 2010 May;42(5):341-7.

(7) Desai A, Sandhu S, Lai JP, Sandhu DS. Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review. World J Hepatol 2019 Jan 27;11(1):1-18.

(8) Geh D, Rana FA, Reeves HL. Weighing the benefits of hepatocellular carcinoma surveillance against potential harms. J Hepatocell Carcinoma 2019;6:23-30.

(9) Assi N, Fages A, Vineis P, Chadeau-Hyam M, Stepien M, Duarte-Salles T, Byrnes G, Boumaza H, Knuppel S, Kuhn T, Palli D, Bamia C, et al. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. Mutagenesis 2015 Nov;30(6):743-53.

(10) Beyoglu D, Idle JR. The metabolomic window into hepatobiliary disease. J Hepatol 2013 Oct;59(4):842-58.

(11) Le ML, Triba MN, Nahon P, Bouchemal N, Hantz E, Goossens C, Amathieu R, Savarin P. Nuclear magnetic resonance metabolomics and human liver

diseases: The principles and evidence associated with protein and carbohydrate metabolism. Biomed Rep 2017 Apr;6(4):387-95.

(12) Huang Q, Tan Y, Yin P, Ye G, Gao P, Lu X, Wang H, Xu G. Metabolic characterization of hepatocellular carcinoma using nontargeted tissue metabolomics. Cancer Res 2013 Aug 15;73(16):4992-5002.

(13) Lu Y, Li N, Gao L, Xu YJ, Huang C, Yu K, Ling Q, Cheng Q, Chen S, Zhu M, Fang J, Chen M, et al. Acetylcarnitine Is a Candidate Diagnostic and Prognostic Biomarker of Hepatocellular Carcinoma. Cancer Res 2016 May 15;76(10):2912-20.

(14) Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, Yin Z, Xie X, Wang X, Chen W, Zhou L, Wang X, et al. A Large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. Hepatology 2018 Feb;67(2):662-75.

(15) Kimhofer T, Fye H, Taylor-Robinson S, Thursz M, Holmes E. Proteomic and metabonomic biomarkers for hepatocellular carcinoma: a comprehensive review. Br J Cancer 2015 Mar 31;112(7):1141-56.

(16) Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, Achaintre D, Assi N, Tjonneland A, Overvad K, Bastide N, Boutron-Ruault MC, et al. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. Int J Cancer 2016 Jan 15;138(2):348-60.

(17) Fages A, Duarte-Salles T, Stepien M, Ferrari P, Fedirko V, Pontoizeau C, Trichopoulou A, Aleksandrova K, Tjonneland A, Olsen A, Clavel-Chapelon F, Boutron-Ruault MC, et al. Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort. BMC Med 2015 Sep 23;13:242.

(18) Loftfield E, Rothwell JA, Sinha R, Keski-Rahkonen P, Robinot N, Albanes D, Weinstein SJ, Derkach A, Sampson J, Scalbert A, Freedman ND. Prospective Investigation of Serum Metabolites, Coffee Drinking, Liver Cancer Incidence, and Liver Disease Mortality. J Natl Cancer Inst 2020 Mar 1;112(3):286-94.

(19) His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, Ferrari P, Romieu I, Onland-Moret NC, Weiderpass E, Dahm CC, Overvad K, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med 2019 Sep 24;17(1):178.

(20) Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, Gicquiau A, Gunter MJ, Ferrari P, Kaaks R, KÃ¼hn T, Floegel A, et al. Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases

and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition. BMC Med 2017 Jul 5;15(1):122.

(21) Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, Gicquiau A, Gunter MJ, Ferrari P, Kaaks R, KÃ¼hn T, Boeing H, et al. Patterns in metabolite profile are associated with risk of more aggressive prostate cancer: A prospective study of 3,057 matched case-control sets from EPIC. Int J Cancer 2020 Feb 1;146(3):720-30.

(22) Carayol M, Leitzmann MF, Ferrari P, Zamora-Ros R, Achaintre D, Stepien M, Schmidt JA, Travis RC, Overvad K, TjÃ¸nneland A, Hansen L, Kaaks R, et al. Blood Metabolic Signatures of Body Mass Index: A Targeted Metabolomics Study in the EPIC Cohort. J Proteome Res 2017 Sep 1;16(9):3137-46.

(23) Schmidt JA, Rinaldi S, Ferrari P, Carayol M, Achaintre D, Scalbert A, Cross AJ, Gunter MJ, Fensom GK, Appleby PN, Key TJ, Travis RC. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. Am J Clin Nutr 2015 Dec;102(6):1518-26.

(24) Moore SC, Playdon MC, Sampson JN, Hoover RN, Trabert B, Matthews CE, Ziegler RG. A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk. J Natl Cancer Inst 2018 Jun 1;110(6):588-97.

(25) Jee SH, Kim M, Kim M, Yoo HJ, Kim H, Jung KJ, Hong S, Lee JH. Metabolomics Profiles of Hepatocellular Carcinoma in a Korean Prospective Cohort: The Korean Cancer Prevention Study-II. Cancer Prev Res (Phila) 2018 May;11(5):303-12.

(26) Yu B, Zanetti KA, Temprosa M, Albanes D, Appel N, Barrera CB, Ben-Shlomo Y, Boerwinkle E, Casas JP, Clish C, Dale C, Dehghan A, et al. The Consortium of Metabolomics Studies (COMETS): Metabolomics in 47 Prospective Cohort Studies. Am J Epidemiol 2019 Jun 1;188(6):991-1012.

(27) Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van OB, Pujos-Guillot E, Verheij E, Wishart D, Wopereis S. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 2009 Dec;5(4):435-58.

(28) Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S6-14.

(29) Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, et al. European

Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002 Dec;5(6B):1113-24.

(30) Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. Biometrics 1984 Mar;40(1):63-75.

(31) Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res 2013 Jan;41(Database issue):D801-D807.

(32) Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. METLIN: a metabolite mass spectral database. Ther Drug Monit 2005 Dec;27(6):747-51.

(33) Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 2007 Sep;3(3):211-21.

(34) Trichopoulos D, Bamia C, Lagiou P, Fedirko V, Trepo E, Jenab M, Pischon T, Nothlings U, Overved K, Tjonneland A, Outzen M, Clavel-Chapelon F, et al. Hepatocellular carcinoma risk factors and disease burden in a European cohort: a nested case-control study. J Natl Cancer Inst 2011 Nov 16;103(22):1686-95.

(35) Stepien M, Fedirko V, Duarte-Salles T, Ferrari P, Freisling H, Trepo E, Trichopoulou A, Bamia C, Weiderpass E, Olsen A, Tjonneland A, Overvad K, et al. Prospective association of liver function biomarkers with development of hepatobiliary cancers. Cancer Epidemiol 2016 Feb;40:179-87.

(36) Satriano L, Lewinska M, Rodrigues PM, Banales JM, Andersen JB. Metabolic rearrangements in primary liver cancers: cause and consequences. Nat Rev Gastroenterol Hepatol 2019 Dec;16(12):748-66.

(37) Leonard SW, Paterson E, Atkinson JK, Ramakrishnan R, Cross CE, Traber MG. Studies in humans using deuterium-labeled alpha- and gamma-tocopherols demonstrate faster plasma gamma-tocopherol disappearance and greater gamma-metabolite production. Free Radic Biol Med 2005 Apr 1;38(7):857-66.

(38) Jiang Q, Christen S, Shigenaga MK, Ames BN. gamma-tocopherol, the major form of vitamin E in the US diet, deserves more attention. Am J Clin Nutr 2001 Dec;74(6):714-22.

(39) Uray IP, Dmitrovsky E, Brown PH. Retinoids and rexinoids in cancer prevention: from laboratory to clinic. Semin Oncol 2016 Feb;43(1):49-64.

(40) Lai GY, Weinstein SJ, Albanes D, Taylor PR, Virtamo J, McGlynn KA, Freedman ND. Association of serum alpha-tocopherol, beta-carotene, and retinol with liver cancer incidence and chronic liver disease mortality. Br J Cancer 2014 Nov 25;111(11):2163-71.

(41) Yuan JM, Gao YT, Ong CN, Ross RK, Yu MC. Prediagnostic level of serum retinol in relation to reduced risk of hepatocellular carcinoma. J Natl Cancer Inst 2006 Apr 5;98(7):482-90.

(42) Galli F, Piroddi M, Lannone A, Pagliarani S, Tomasi A, Floridi A. A comparison between the antioxidant and peroxynitrite-scavenging functions of the vitamin E metabolites alpha- and gamma-carboxyethyl-6-hydroxychromans. Int J Vitam Nutr Res 2004 Sep;74(5):362-73.

(43) Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM, Bass NM, Neuschwander-Tetri BA, Lavine JE, Tonascia J, Unalp A, Van NM, Clark J, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med 2010 May 6;362(18):1675-85.

(44) Lasky-Su J, Dahlin A, Litonjua AA, Rogers AJ, McGeachie MJ, Baron RM, Gazourian L, Barragan-Bradford D, Fredenburgh LE, Choi AMK, Mogensen KM, Quraishi SA, et al. Metabolome alterations in severe critical illness and vitamin D status. Crit Care 2017 Jul 28;21(1):193.

(45) Fedirko V, Duarte-Salles T, Bamia C, Trichopoulou A, Aleksandrova K, Trichopoulos D, Trepo E, Tjonneland A, Olsen A, Overvad K, Boutron-Ruault MC, Clavel-Chapelon F, et al. Prediagnostic circulating vitamin D levels and risk of hepatocellular carcinoma in European populations: a nested case-control study. Hepatology 2014 Oct;60(4):1222-30.

(46) Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, Yin Z, Xie X, Wang X, Chen W, Zhou L, Wang X, et al. A large-scale, multi-center serum metabolite biomarkers identification study for the early detection of hepatocellular carcinoma. Hepatology 2017 Sep 28.

(47) Barcelos RP, Stefanello ST, Mauriz JL, Gonzalez-Gallego J, Soares FA. Creatine and the Liver: Metabolism and Possible Interactions. Mini Rev Med Chem 2016;16(1):12-8.

(48) Tian YE, Xie XU, Lin Y, Tan G, Zhong WU. Androgen receptor in hepatocarcinogenesis: Recent developments and perspectives. Oncol Lett 2015 May;9(5):1983-8.

(49) Kanda T, Yokosuka O. The androgen receptor as an emerging target in hepatocellular carcinoma. J Hepatocell Carcinoma 2015;2:91-9.

(50) De MN, Manno M, Villa E. Sex hormones and liver cancer. Mol Cell Endocrinol 2002 Jul 31;193(1-2):59-63.

(51) Park CB, Kim DJ, Moore MA, Takasuka N, Tsuda H. Promotion of Liver Lesion Development in the Syrian Hamster by Dietary Fat Following Multi-organ Initiation is Inhibited by DHEA-S Administration. Asian Pac J Cancer Prev 2000;1(4):329-32.

(52) Duarte-Salles T, Fedirko V, Stepien M, Trichopoulou A, Bamia C, Lagiou P, Lukanova A, Trepo E, Overvad K, Tjonneland A, Halkjaer J, Boutron-Ruault MC, et al. Dairy products and risk of hepatocellular carcinoma: the European Prospective Investigation into Cancer and Nutrition. Int J Cancer 2014 Oct 1;135(7):1662-72.

(53) Ishikawa T. Branched-chain amino acids to tyrosine ratio value as a potential prognostic factor for hepatocellular carcinoma. World J Gastroenterol 2012 May 7;18(17):2005-8.

(54) Rauschenbach MO, Zharova EI, Sergeeva TI, Ivanova VD, Probatova NA. Blastomogenic activity of p-hydroxyphenyllactic acid in mice. Cancer Res 1975 Mar;35(3):577-85.

(55) Yang Y, Liu F, Wan Y. Simultaneous determination of 4-hydroxyphenyl lactic acid, 4-hydroxyphenyl acetic acid, and 3,4-hydroxyphenyl propionic acid in human urine by ultra-high performance liquid chromatography with fluorescence detection. J Sep Sci 2017 May;40(10):2117-22.

(56) Medvedev A, Buneeva O, Glover V. Biological targets for isatin and its analogues: Implications for therapy. Biologics 2007 Jun;1(2):151-62.

(57) Fedirko V, Tran HQ, Gewirtz AT, Stepien M, Trichopoulou A, Aleksandrova K, Olsen A, Tjonneland A, Overvad K, Carbonnel F, Boutron-Ruault MC, Severi G, et al. Exposure to bacterial products lipopolysaccharide and flagellin and hepatocellular carcinoma: a nested case-control study. BMC Med 2017 Apr 4;15(1):72.

(58) Koch M, Freitag-Wolf S, Schlesinger S, Borggrefe J, Hov JR, Jensen MK, Pick J, Markus MRP, Hopfner T, Jacobs G, Siegert S, Artati A, et al. Serum metabolomic profiling highlights pathways associated with liver fat content in a general population sample. Eur J Clin Nutr 2017 Aug;71(8):995-1001.

(59) Sugimoto H, Sakurai S, Abe T, Takagi H, Takahashi H, Takezawa J, Nagamine T, Matsuzaki S. Elevation of N1-acetylspermidine and putrescine in hepatic tissues of patients with fulminant hepatitis and liver cirrhosis. J Gastroenterol 1994 Apr;29(2):159-63.

(60) Xu H, Liu R, He B, Bi CW, Bi K, Li Q. Polyamine Metabolites Profiling for Characterization of Lung and Liver Cancer Using an LC-Tandem MS Method with Multiple Statistical Data Mining Strategies: Discovering Potential Cancer Biomarkers in Human Plasma and Urine. Molecules 2016 Aug 10;21(8).

(61) Grammatikos G, Muhle C, Ferreiros N, Schroeter S, Bogdanou D, Schwalm S, Hintereder G, Kornhuber J, Zeuzem S, Sarrazin C, Pfeilschifter J. Serum acid sphingomyelinase is upregulated in chronic hepatitis C infection and non alcoholic fatty liver disease. Biochim Biophys Acta 2014 Jul;1841(7):1012-20.

(62) Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. Arch Intern Med 1999 Dec 13;159(22):2647-58.

(63) Trottier J, Bialek A, Caron P, Straka RJ, Heathcote J, Milkiewicz P, Barbier O. Metabolomic profiling of 17 bile acids in serum from patients with primary biliary cirrhosis and primary sclerosing cholangitis: a pilot study. Dig Liver Dis 2012 Apr;44(4):303-10.

(64) Safaei A, Arefi OA, Mohebbi SR, Rezaei-Tavirani M, Mahboubi M, Peyvandi M, Okhovatian F, Zamanian-Azodi M. Metabolomic analysis of human cirrhosis, hepatocellular carcinoma, non-alcoholic fatty liver disease and non-alcoholic steatohepatitis diseases. Gastroenterol Hepatol Bed Bench 2016;9(3):158-73.

(65) Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M, Honda K, Ishikawa Y, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 2013 Jul 4;499(7456):97-101.

(66) Chen T, Xie G, Wang X, Fan J, Qiu Y, Zheng X, Qi X, Cao Y, Su M, Wang X, Xu LX, Yen Y, et al. Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. Mol Cell Proteomics 2011 Jul;10(7):M110.

(67) Tan Y, Yin P, Tang L, Xing W, Huang Q, Cao D, Zhao X, Wang W, Lu X, Xu Z, Wang H, Xu G. Metabolomics study of stepwise hepatocarcinogenesis from the model rats to patients: potential biomarkers effective for small hepatocellular carcinoma diagnosis. Mol Cell Proteomics 2012 Feb;11(2):M111.

(68) Chao MR, Wang CJ, Yang HH, Chang LW, Hu CW. Rapid and sensitive quantification of urinary N7-methylguanine by isotope-dilution liquid chromatography/electrospray ionization tandem mass spectrometry with on-line solid-phase extraction. Rapid Commun Mass Spectrom 2005;19(17):2427-32.

(69) Tamae K, Kawai K, Yamasaki S, Kawanami K, Ikeda M, Takahashi K, Miyamoto T, Kato N, Kasai H. Effect of age, smoking and other lifestyle factors on urinary 7-methylguanine and 8-hydroxydeoxyguanosine. Cancer Sci 2009 Apr;100(4):715-21.

(70) Romaguera D, Vergnaud AC, Peeters PH, van Gils CH, Chan DS, Ferrari P, Romieu I, Jenab M, Slimani N, Clavel-Chapelon F, Fagherazzi G, Perquier F, et al. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. Am J Clin Nutr 2012 Jul;96(1):150-63.

(71) Huang J, Weinstein SJ, Moore SC, Derkach A, Hua X, Liao LM, Gu F, Mondul AM, Sampson JN, Albanes D. Serum Metabolomic Profiling of All-Cause Mortality: A Prospective Analysis in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study Cohort. Am J Epidemiol 2018 Jan 30.

(72) Ivanisevic J, Zhu ZJ, Plate L, Tautenhahn R, Chen S, O'Brien PJ, Johnson CH, Marletta MA, Patti GJ, Siuzdak G. Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism. Anal Chem 2013 Jul 16;85(14):6876-84.

Table Titles

Table 1. Characteristics of Hepatocellular Cancer (HCC) cases and matched control subjects, nested within the EPIC cohort.

Table 2. Associations with risk of Hepatocellular Carcinoma (HCC) development for Level 1* identified metabolites.

Table 3. Associations with risk of Hepatocellular Carcinoma (HCC) development for Level 2* (33) identified metabolites.

Table 4. Associations with risk of Hepatocellular Carcinoma (HCC) development for Level 3 identified and Level 4 unidentified metabolites*.

Figure Titles and Legends

Figure 1 Title: Flow chart of the selection procedures for metabolites and number of annotated compounds for each analytical configuration of the UHPLC-QTOF-MS system.

Figure 1 Legend: A total of 114 separate compounds (i.e. confirmed molecules that consisted of one or more features) were identified from the four datasets. Of these 114 separate compounds, 22 were also detected by more than at least one of the other three profiling methods, leaving a total of 92 unique compounds. Of these 92 compounds, 46 were identified into 3 distinct categories: unambiguously identified using pure standards (Level 1; n=14), identified to a high level of confidence based on chemical features and characteristics (Level 2; n=23), and identified to a known

chemical class (Level 3; n=9). The remaining 46 compounds were not identified, i.e. unknown.

[1] After Benjamini-Hochberg correction for multiple testing, conditioned on matching factors: age at blood collection (±1 year), sex, study center, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6,and >6 hours); among women, additionally by menopausal status (pre-, peri-, and postmenopausal), and hormone replacement therapy use at time of blood collection (yes/no).

[2] After Benjamini-Hochberg correction for multiple testing: matching factors + BMI (kg/m2, continuous), waist circumference (cm, continuous), physical activity (Met-h/wk, continuous), alcohol intake at recruitment (g/d, continuous), lifetime alcohol intake pattern (categorical), smoking status (categorical) and attained education (categorical).

Please see Tables 2 to 4 for additional details.

Figure 2 Title: Principal component (PC) analyses based on metabolites associated with HCC risk for **(1)** the 46 metabolites associated with HCC risk and identified at Levels 1 to 3 [33] and **(2)** the 14 metabolites associated with HCC risk and identified at Level 1 only, i.e. unambiguous identification using pure standards[33]. HCC cases are shown by green circles and matched controls by mauve triangles.

Figure 2 Legend: (A) Score plots of PC analyses differentiating cases and controls, **(B)** plot of scores on PC1 versus follow-up time (all years, number of HCC case and matched control sets=129; and excluding cases with 4 or less years of follow-up,

number of HCC case and matched control sets=87) and **(C)** relative contributions of identified metabolites to PC1 and PC2. For **A1**, the proportion of variability is 29.86% for PC1 vs 11.12% for PC2. For **A2**, the values are 29.18% for PC1 and 11.16% for PC2.

Novelty and Impact:

Changes in liver function precede the development of hepatocellular carcinoma (HCC). Many of these changes can be detected in the blood, as can biomarkers related to lifestyle or environmental exposures that may affect HCC risk. In this study, based on a large, prospective observational cohort, the authors used high resolution mass spectrometry-based metabolomics to identify alterations in circulating levels of 92 metabolites associated with HCC risk, 14 of which could be annotated with high confidence and some of which were observed up to 10 years prior to diagnosis. These results offer insight into early metabolic perturbations and mechanisms leading to this deadly cancer.

TABLES

Table 1. Characteristics of Hepatocellular Cancer (HCC) cases and matched control subjects, nested within the EPIC cohort.

Variables		HCC Cases n=129		Matched Controls n=129		p value*
Women (n, %)		41	31.8	41	31.8	
Age at recruitment (years), mean (SD)		60.0	7.3	60.1	7.4	0.717
BMI (kg/m²), mean (SD)		28.4	4.6	27.4	4.3	0.062
Waist circumference (cm), mean (SD)		97.5	14.0	93.0	12.3	0.002
Physical activity (MET-h/week), mean (SD)		83.4	54.2	85.0	50.8	0.873
Dietary alcohol (g/day), mean (SD)		22.5	35.9	15.8	20.0	0.058
Education (n, %)	None / primary	68	52.7	64	49.2	0.5945
	Technical / professional	34	26.4	29	22.3	
	Secondary	6	4.7	10	7.8	
	University or higher	19	14.7	23	17.8	
Alcohol intake pattern (n, %)	Never drinkers	9	7.0	12	9.3	0.0018
	Former drinkers	23	17.8	4	3.1	
	Drinkers only at recruitment	7	5.4	8	6.2	
	Always drinkers	90	69.8	105	80.8	
Smoking status (n, %)	Never smokers	42	32.6	60	46.5	0.0124
	Former smokers	40	31.0	42	32.3	
	Current smokers	46	35.7	26	20.0	
Hepatitis B and/or C infection (n, %) **	**Yes**	33	25.6	4	3.1	<0.0001

HCV (n, %)	**Yes**	19	14.7	2	1.5	<0.0001
HBV (n, %)	**Yes**	17	13.2	3	2.3	<0.0001
Self-reported diabetes status at baseline (n, %)	**Yes**	12	9.3	6	4.7	0.1426
Liver function score (n, %) ≠	**0**	24	18.6	75	58.1	<0.0001
	≥1	66	51.2	16	12.4	

Missing values were not excluded from percentage calculations, thus the sum of percent values across sub-groups may not add up to 100%.

Number of cases and controls with missing or unknown variable value: education (controls=3, HHCC=2), smoking status (controls=1, HCC=1), hepatitis infection status (controls=38, HC =38), diabetes status (self-reported, controls=13, HCC =11), liver function score (controls=38, HCC=38). The distribution of cases by country is as follows: Denmark=23, Germany=32, Greece=16, Italy=28, the Netherlands=4, Spain=11, United Kingdom=15.

Categorical variables are presented as numbers and percentages. Continuous variables are presented as mean and standard deviation (SD).

Paired t-test for continuous and Fisher's exact test for categorical variables were used to calculate p-value.

** Hepatitis B and/or C seropositivity were detected using the ARCHITECT HBsAg and anti-HCV chemiluminescent microparticle immunoassays (CMIAs; Abbott Diagnostics, France)

≠ Liver function biomarkers (ALT, AST, GGT, ALP, albumin, bilirubin) were measured on the ARCHITECT c Systems™ (Abbott Diagnostics). A liver function score was computed as an indicator of possible underlying liver damage. The score ranges from 0 to 6 and is based on abnormal liver function tests (ALT>55 U/L, AST>34 U/L, GGT >64 U/L for men and > 36 U/L for women, ALP > 150 U/L, albumin < 34 g/L, total bilirubin > 20.5 µmol/L; values were provided by the laboratory). For each liver function biomarker, participants with abnormal values (as defined above) were assigned a score of 1. Possible liver impairment category was created for the score ≥1.

Table 2. Associations with risk of Hepatocellular Carcinoma (HCC) development for Level 1* identified metabolites.

Identified and Annotated Compound [1]	Analytical Method	*m/z*	Retention Time (mins)	Absolute Fold Change[2]	Multivariable Adjusted [3] OR (95% CI)	q-value
Retinol	RP+	269.2273	7.27	-1.30	0.27 (0.16 - 0.48)	0.00060
Dehydroepiandrosterone Sulfate**	HILIC-	367.1561	0.80	-2.13	0.35 (0.22 - 0.57)	0.00350
Glycerophosphocholine	RP+	280.0920	0.64	-1.47	0.44 (0.28 - 0.71)	0.01080
γ-carboxyethyl hydroxychroman	RP+	265.1428	5.31	-1.23	0.56 (0.39 - 0.81)	0.01970
Creatine	RP+	132.0771	0.66	-1.20	0.56 (0.37 - 0.83)	0.03410
N1-Acetylspermidine* *	HILIC+	188.1759	6.85	1.20	2.16 (1.38 - 3.37)	0.01370
Isatin	RP+	148.0393	3.35	1.39	2.56 (1.53 – 4.29)	0.01490
p-Hydroxyphenyllactic acid	HILIC-	181.0494	2.24	1.47	2.63 (1.62 – 4.28)	0.02200
Tyrosine	RP+	182.0816	1.28	1.20	2.77 (1.58 - 4.83)	0.02030
Sphingosine	RP+	300.2902	6.06	1.36	2.79 (1.66 - 4.71)	0.00360
L,L-Cyclo(leucylprolyl)	RP+	211.1442	3.90	2.37	3.25 (1.91 - 5.53)	0.00080

Glycochenodeoxycholic acid **	RP+	450.3218	6.48	3.37	3.31 (1.99 - 5.51)	0.00050
Glycocholic acid **	RP+	466.3164	6.21	3.92	4.07 (2.32 - 7.14)	0.00040
7-methylguanine	HILIC+	166.0729	2.53	1.31	6.78 (3.24 - 14.18)	0.00030

* features identified with high confidence and verified by a chemical standard

** indicates that a compound was detected by more than one method. The listed method is the one showing the greatest intensity for the particular compound. Dehydroepiandrosterone Sulfate was also detected by RP-; N1-Acetylspermidine was also detected by RP+; Glycochenodeoxycholic acid was also detected by HILIC+, RP-; Glycocholic acid was also detected by RP-; Benzylcarnitine was also detected by RP+.

[1] Level 1 identified compounds: retention time and MS/MS matches with an authentic chemical standard (33). Information for compounds identified at Level 2 (identified compounds with high confidence; no standard available/analysed but matching isotope pattern, MS/MS spectra, and other supporting evidence) is shown on **Table 3**. Information for compounds identified at Level 3 (compounds identified from a known chemical class) and Level 4 (unidentified compounds) are shown in **Table 4**.

[2] Absolute fold change between the median intensities of cases to their matched controls.

[3] The ORs represent the risk of HCC per 1 standard deviation (SD) of logarithm transformed value. Multivariable adjusted: matching factors + body mass index (BMI, kg/m2, continuous), Waist circumference (cm, continuous), alcohol intake at recruitment (g/d, continuous), Physical activity (Met-h week, continuous), categories of smoking status, alcohol intake pattern and education (for categories see Table 1).

Table 3. Associations with risk of Hepatocellular Carcinoma (HCC) development for Level 2* identified metabolites.

Identified and Annotated Compound [1]	Analytical Method	*m/z*	Retention Time (mins)	Absolute Fold Change[2]	Multivariable Adjusted [3] OR (95% CI)	q-value
LysoPC(17:0) **	HILIC+	510.3554	2.42	-1.64	0.21 (0.11 - 0.40)	0.00030
LysoPC(15:0) **	HILIC+	482.3251	2.60	-1.54	0.23 (0.12 - 0.42)	0.00030
LysoPC(20:5) **	HILIC+	564.3104	2.38	-1.52	0.23 (0.12 - 0.46)	0.00030
LysoPC(16:0) **	RP+*	991.6772	7.03	-1.26	0.28 (0.16 - 0.49)	0.00060
LysoPC(20:4) **	HILIC+	544.3407	2.33	-1.40	0.31 (0.19 - 0.51)	0.00030
LysoPC(P-16:0) **	HILIC+	480.3460	2.12	-1.25	0.33 (0.19 - 0.55)	0.00030
LysoPC(22:5)	HILIC+	570.3539	2.30	-1.33	0.33 (0.20 - 0.54)	0.00030
PC(38:6)	RP+	806.5690	8.51	-1.29	0.36 (0.21 - 0.61)	0.00460
LysoPC(22:6) **	HILIC+	568.3399	2.28	-1.41	0.37 (0.23 - 0.58)	0.00030
LysoPC(18:2) **	HILIC+	520.3418	2.46	-1.31	0.40 (0.26 - 0.64)	0.00030
LysoPC(18:0) **	HILIC+	524.3712	2.34	-1.25	0.41 (0.26 - 0.65)	0.00250
C5 acylcarnitine	HILIC+	246.1703	3.17	-1.22	0.46 (0.29 - 0.73)	0.01770
DG(18:2/18:2/0:0)	RP+	639.4946	9.54	-1.39	0.47 (0.31 - 0.72)	0.00950
LysoPC(14:0) **	HILIC+	468.3088	2.70	-1.26	0.48 (0.31 - 0.72)	0.00900
LysoPC(17:1)	HILIC+	508.3406	2.46	-1.27	0.48 (0.31 - 0.75)	0.01530
LysoPC (18:1)	HILIC+	522.3570	2.38	-1.24	0.52 (0.34 - 0.78)	0.02450
LysoPC(20:3) **	RP+	546.3548	7.04	-1.21	0.56 (0.39 - 0.80)	0.01790
DG(18:1/18:2/0:0)	RP+	641.5106	10.10	-1.27	0.58 (0.40 - 0.85)	0.04060
PC(16:1/16:1/0:0)	HILIC+	730.5398	1.17	1.39	1.79 (1.22 - 2.64)	0.03570
Bilirubin isomer 2	RP+	585.2687	4.34	1.26	1.89 (1.20 - 2.97)	0.04510
Bilirubin isomer 1	RP+	585.2696	5.13	1.27	1.94 (1.22 - 3.06)	0.03850
PC(16:1/16:0/0:0) **	HILIC+	732.5552	1.16	1.54	2.01 (1.33 - 3.03)	0.01410

Benzoylcarnitine **	HILIC+	266.1392	3.36	1.46	2.74 (1.69 - 4.42)	0.00030

- features identified with high confidence

** indicates that a compound was detected by more than one method. The listed method is the one showing the greatest intensity for the particular compound. LysoPC(17:0) was also detected by RP+/-; LysoPC(15:0) also by RP+; LysoPC(20:5) also by HILIC-, RP+/-; LysoPC(16:0) also by HILIC-; LysoPC(20:4) also by HILIC-, RP+/-; LysoPC(P-16:0) also by RP+; LysoPC(22:6) also by RP+/-; LysoPC(18:2) also by RP+, HILIC-; LysoPC(18:0) also by RP+, HILIC-; LysoPC(14:0) also by RP+; LysoPC(20:3) also by HILIC-; PC(16:1/16:0/0:0) also by RP+; Benzoylcarnitine also by RP+.

[1] Level 1 identified compounds: retention time and MS/MS matches with an authentic chemical standard (**Table 2**); Level 2 (identified compounds with high confidence): no standard available/analysed but matching isotope pattern, MS/MS spectra, and other supporting evidence (33). Information for compounds identified at Level 3 (compounds identified from a known chemical class) and Level 4 (unidentified compounds) are shown in **Table 4**.

[2] Absolute fold changes between the median intensities of cases to their matched controls.

[3] The OR represents the risk of HCC per 1 SD of logarithm transformed value. Multivariable adjusted: matching factors + body mass index (BMI, kg/m2, continuous), Waist circumference (cm, continuous), alcohol intake at recruitment (g/d, continuous), Physical activity (Met-h wk, continuous), categories of smoking status, alcohol intake pattern and education (for categories see Table 1).

Table 4. Associations with risk of Hepatocellular Carcinoma (HCC) development for Level 3 identified and Level 4 unidentified metabolites*.

Level of Identification [1]	Chemical Class of Compound	Analytical Method	*m/z*	Retention Time (min)	Absolute Fold Change[2]	Multivariable Adjusted [3] OR (95% CI)	q-value
3	C19H30O2-sulfate (Steroid-S)	HILIC-	369.1710	0.81	-3.79	0.20 (0.11 - 0.39)	0.00080
3	Leucyl-Valine or isomer	RP+	231.1703	2.28	-2.32	0.28 (0.16 - 0.49)	0.00050
3	LysoPC/PC	HILIC+	633.3975	2.22	-1.67	0.35 (0.21 - 0.57)	0.00030
3	LysoPC(18:2) isomer	HILIC-	564.3263	2.09	-1.30	0.39 (0.24 - 0.63)	0.01000
3	LysoPC/PC	HILIC+	609.3992	2.24	-1.38	0.41 (0.26 - 0.63)	0.00030
3	Tryptophyl-phenylalanine	RP+	352.1663	3.61	-1.30	0.48 (0.30 - 0.76)	0.01750
3	L,L-Cyclo(isoleucylprolyl)	RP+	211.1443	3.79	1.45	1.86 (1.21 - 2.84)	0.03600
3	C19H30O3-sulfate (OH-Steroid-S)	RP-	385.1661	5.05	1.29	2.30 (1.46 - 3.61)	0.03050
3	C19H28O3-sulfate (OH-DHEA-S)**	RP-	383.1505	5.48	1.87	2.59 (1.67 - 4.01)	0.00410
4	Unknown	RP+	551.3114	6.94	-1.32	0.22 (0.11 - 0.42)	0.00060
4	Unknown	RP+	571.2988	6.93	-1.30	0.24 (0.13 - 0.43)	0.00040

Level of Identification [1]	Chemical Class of Compound	Analytical Method	*m/z*	Retention Time (min)	Absolute Fold Change[2]	Multivariable Adjusted [3] OR (95% CI)	q-value
4	Unknown	RP+	794.9647	6.96	-1.35	0.31 (0.18 - 0.54)	0.00160
4	Unknown	RP+	268.1413	0.88	-1.34	0.33 (0.18 - 0.62)	0.00950
4	Unknown	RP+	239.0915	4.29	-1.76	0.35 (0.19 - 0.63)	0.00910
4	Unknown	RP+	543.3458	7.28	-1.27	0.39 (0.24 - 0.62)	0.00330
4	Unknown	RP+	203.1392	0.88	-1.52	0.39 (0.23 - 0.68)	0.01160
4	Unknown	RP+	169.9858	0.62	-1.34	0.41(0.23 - 0.71)	0.01680
4	Unknown	RP+	203.1391	1.65	-1.46	0.42 (0.24 - 0.72)	0.01850
4	Unknown	RP+	548.3020	6.95	-1.29	0.42 (0.27 - 0.66)	0.00480
4	Unknown	RP+	500.2774	6.81	-1.25	0.44 (0.27 - 0.70)	0.00940
4	Unknown	RP+	257.2267	6.91	-1.49	0.45 (0.29 - 0.70)	0.00770
4	Unknown	HILIC+	116.1064	1.69	-1.38	0.46 (0.30 - 0.69)	0.00390
4	Unknown	RP+	423.7686	8.82	-1.30	0.46 (0.30 - 0.71)	0.00850
4	Unknown	RP-	228.9786	3.40	-1.31	0.47 (0.31 - 0.72)	0.03380
4	Unknown	RP+	536.3023	6.97	-1.29	0.47 (0.30 - 0.73)	0.01090
4	Unknown	RP+	283.1552	6.21	-1.36	0.48 (0.30 - 0.75)	0.01680
4	Unknown	RP+	541.3301	7.10	-1.20	0.48 (0.30 - 0.76)	0.01880

Level of Identification [1]	Chemical Class of Compound	Analytical Method	*m/z*	Retention Time (min)	Absolute Fold Change[2]	Multivariable Adjusted [3] OR (95% CI)	q-value
4	Unknown	RP+	401.3414	7.72	-1.21	0.53 (0.35 - 0.80)	0.02470
4	Unknown	HILIC+	183.1120	1.31	-1.30	0.54 (0.37 - 0.80)	0.02630
4	Unknown	RP+	118.0498	0.86	-1.21	0.55 (0.36 - 0.82)	0.02950
4	Unknown	RP+	254.0234	4.23	1.51	1.62 (1.17 - 2.24)	0.03190
4	Unknown	RP+	330.2464	5.98	1.22	1.78 (1.22 - 2.61)	0.02690
4	Unknown	RP+	243.1954	6.71	1.40	1.81 (1.24 - 2.63)	0.02020
4	Unknown	RP+	281.2489	6.81	1.42	1.82 (1.24 - 2.67)	0.02250
4	Unknown	RP+	175.0264	1.66	1.57	1.88 (1.23 - 2.87)	0.03130
4	Unknown	HILIC+	120.0657	1.83	1.22	1.97 (1.23 - 3.16)	0.04940
4	Unknown	RP+	241.1543	3.78	1.30	2.00 (1.30 - 3.09)	0.01880
4	Unknown	RP+	104.0710	0.64	1.27	2.05 (1.37 - 3.06)	0.00910
4	Unknown	RP+	202.1187	0.87	1.34	2.06 (1.30 - 3.25)	0.02070
4	Unknown	HILIC-	308.0712	2.34	1.40	2.17 (1.42 - 3.31)	0.02200
4	Unknown	RP+	203.1393	0.73	1.27	2.20 (1.41 - 3.45)	0.00950
4	Unknown	RP+	129.0649	1.59	1.23	2.26 (1.42 - 3.59)	0.00990

Level of Identification [1]	Chemical Class of Compound	Analytical Method	*m/z*	Retention Time (min)	Absolute Fold Change[2]	Multivariable Adjusted [3] OR (95% CI)	q-value
4	Unknown	RP+	129.0661	0.65	1.28	2.34 (1.45 - 3.74)	0.00840
4	Unknown	RP-	71.0501	2.78	1.23	2.35 (1.44 - 3.82)	0.03690
4	Unknown	RP-	475.3034	6.86	1.41	2.37 (1.57 - 3.60)	0.00680
4	Unknown	RP+	163.0752	2.09	1.20	2.38 (1.46 - 3.88)	0.00890
4	Unknown	RP+	203.0214	2.78	1.20	2.46 (1.43 - 4.24)	0.01470
4	Unknown	HILIC+	203.1395	5.32	1.37	2.51 (1.55 - 4.05)	0.00390
4	Unknown	RP-	146.0448	0.65	1.34	2.56 (1.53 - 4.29)	0.03120
4	Unknown	RP+	619.5268	7.00	1.49	2.57 (1.61 - 4.13)	0.00280
4	Unknown	RP+	182.0814	0.87	1.25	2.63 (1.62 - 4.28)	0.00290
4	Unknown	HILIC+	126.0662	3.35	1.20	2.67 (1.59 - 4.48)	0.00390
4	Unknown	RP+	431.3169	6.85	1.38	2.80 (1.76 - 4.47)	0.00090
4	Unknown	RP+	389.2650	6.37	1.77	3.22 (1.77 - 5.85)	0.00380
4	Unknown	RP+	614.5721	7.00	1.25	3.75 (1.99 - 7.05)	0.00190

* features that are identified at the level of the chemical class (Level 3) or unknown (Level 4) (33).

[1] indicates that a compound was detected by more than one method, also listed. Data are provided only for the method that showed the greatest intensity for the particular compound: C19H28O3-sulfate (OH-DHEA-S) was also detected by HILIC-ve.

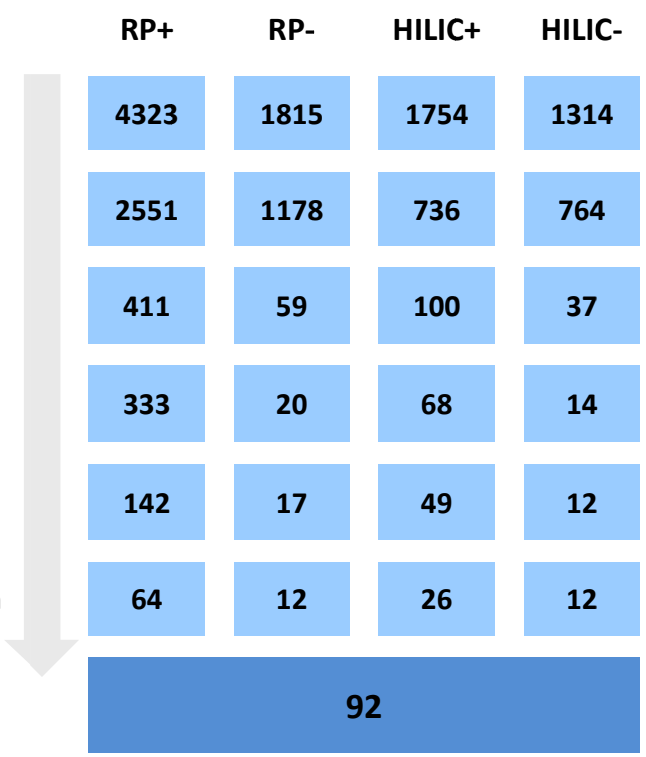
[1] Compounds identified at Level 3 (compound from a known chemical class) and Level 4 (unknown compounds) are shown here (33). Information for compounds identified at Level 1 (retention time and MS/MS matches with an authentic chemical standard) and Level 2 (no standard available/analysed but matching isotope pattern, MS/MS spectra, and other supporting evidence) are shown in **Tables 2 and 3, respectively**.

[2] Absolute fold change between the median intensities of cases to their matched controls.

[3] The OR represent the risk of HCC per 1 SD of logarithm transformed value. Multivariable adjusted: matching factors + body mass index (BMI, kg/m2, continuous), Waist circumference (cm, continuous), alcohol intake at recruitment (g/d, continuous), Physical activity (Met-h wk, continuous), categories of smoking status, alcohol intake pattern and education (for categories see Table 1).

LysoPC=lysophosphatidylcholine; Unknown = not identifiable; identity or chemical class not ascertainable.

Figure 1: Flow chart of the selection procedures for metabolites and number of annotated compounds for each analytical configuration of the UHPLC-QTOF-MS system.



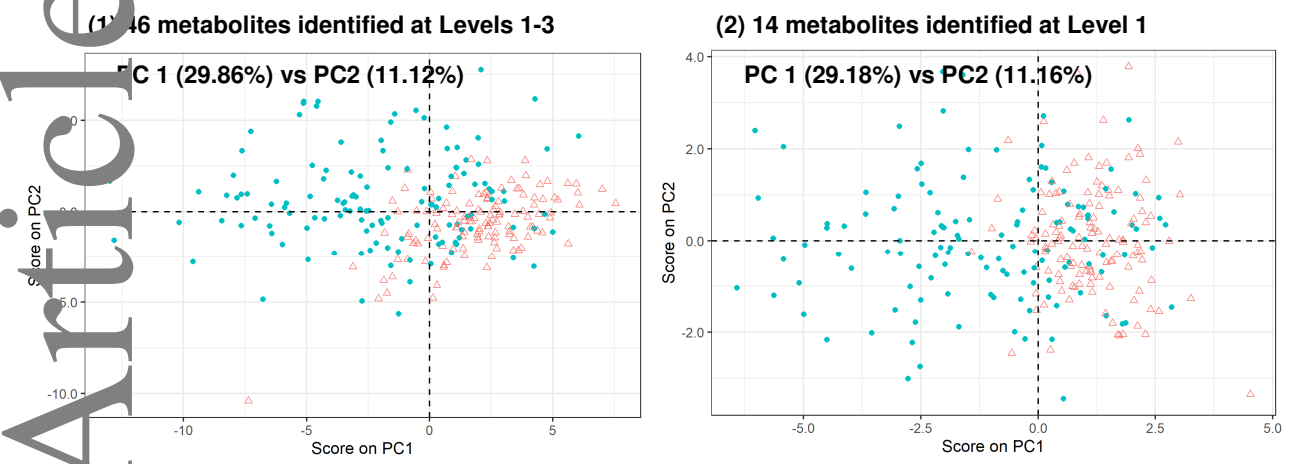
A total of 114 separate compounds (i.e. confirmed molecules that consisted of one or more features) were identified from the four datasets. Of these 114 separate compounds, 22 were also detected by more than at least one of the other three profiling methods, leaving a total of 92 unique compounds. Of these 92 compounds, 46 were identified into 3 distinct categories: unambiguously identified using pure standards (Level 1; n=14), identified to a high level of confidence based on chemical features and characteristics (Level 2; n=23), and identified to a known chemical class (Level 3; n=9). The remaining 46 compounds were not identified, i.e. unknown.

[1] After Benjamini-Hochberg correction for multiple testing, conditioned on matching factors: age at blood collection (±1 year), sex, study center, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6,and >6 hours); among women, additionally by menopausal status (pre-, peri-, and postmenopausal), and hormone replacement therapy use at time of blood collection (yes/no).
[2] After Benjamini-Hochberg correction for multiple testing: matching factors + BMI (kg/m2, continuous), waist circumference (cm, continuous), physical activity (Met-h/wk, continuous), alcohol intake at recruitment (g/d, continuous), lifetime alcohol intake pattern (categorical), smoking status (categorical) and attained education (categorical).

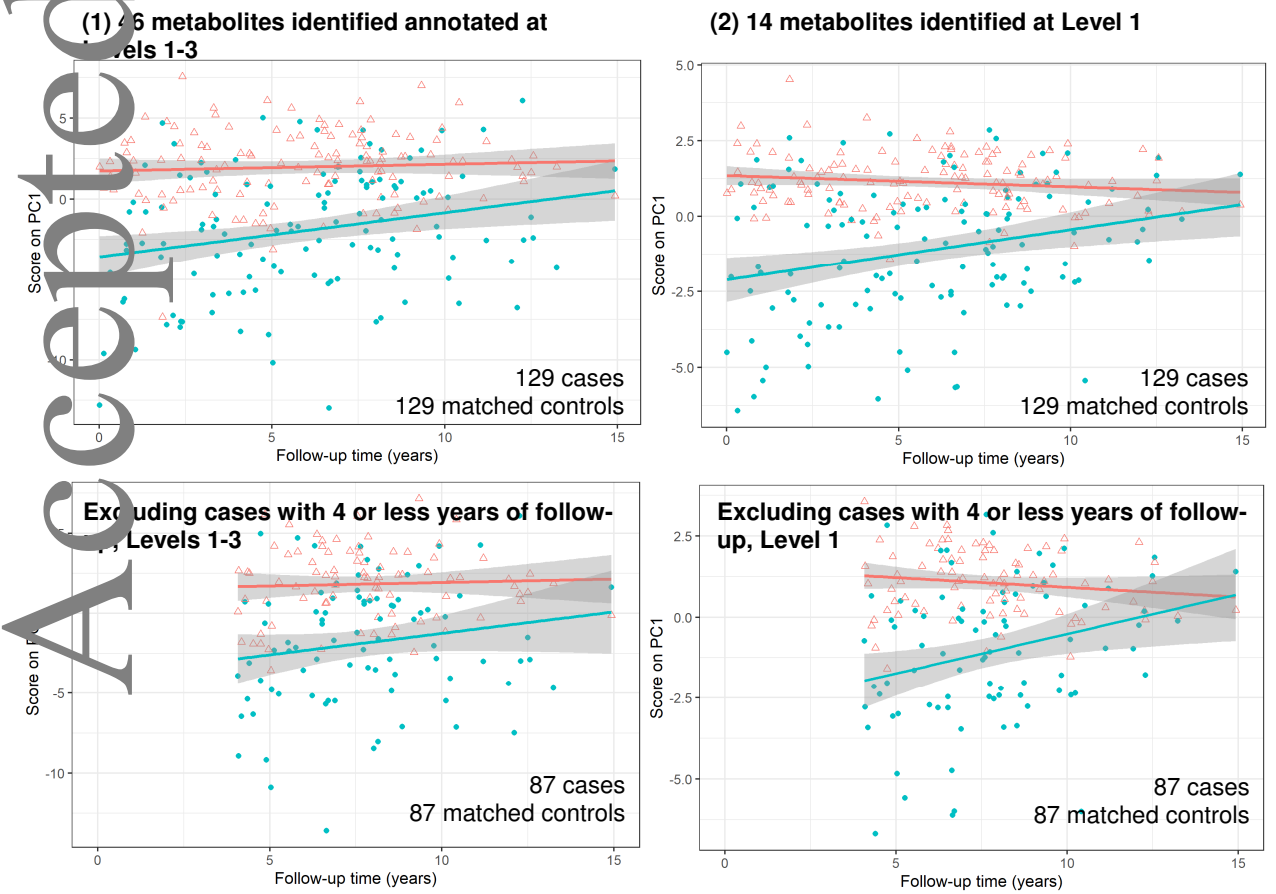
Please see Tables 2 to 4 for additional details.

Figure 2. Principal component (PC) analyses based on metabolites associated with HCC risk for **(1)** the 46 metabolites associated with HCC risk and identified at Levels 1 to 3 [33] and **(2)** the 14 metabolites associated with HCC risk and identified at Level 1 only, i.e. unambiguous identification using pure standards[33]. HCC cases are shown by green circles and matched controls by mauve triangles.

A. Score plots of PC analyses differentiating cases and controls

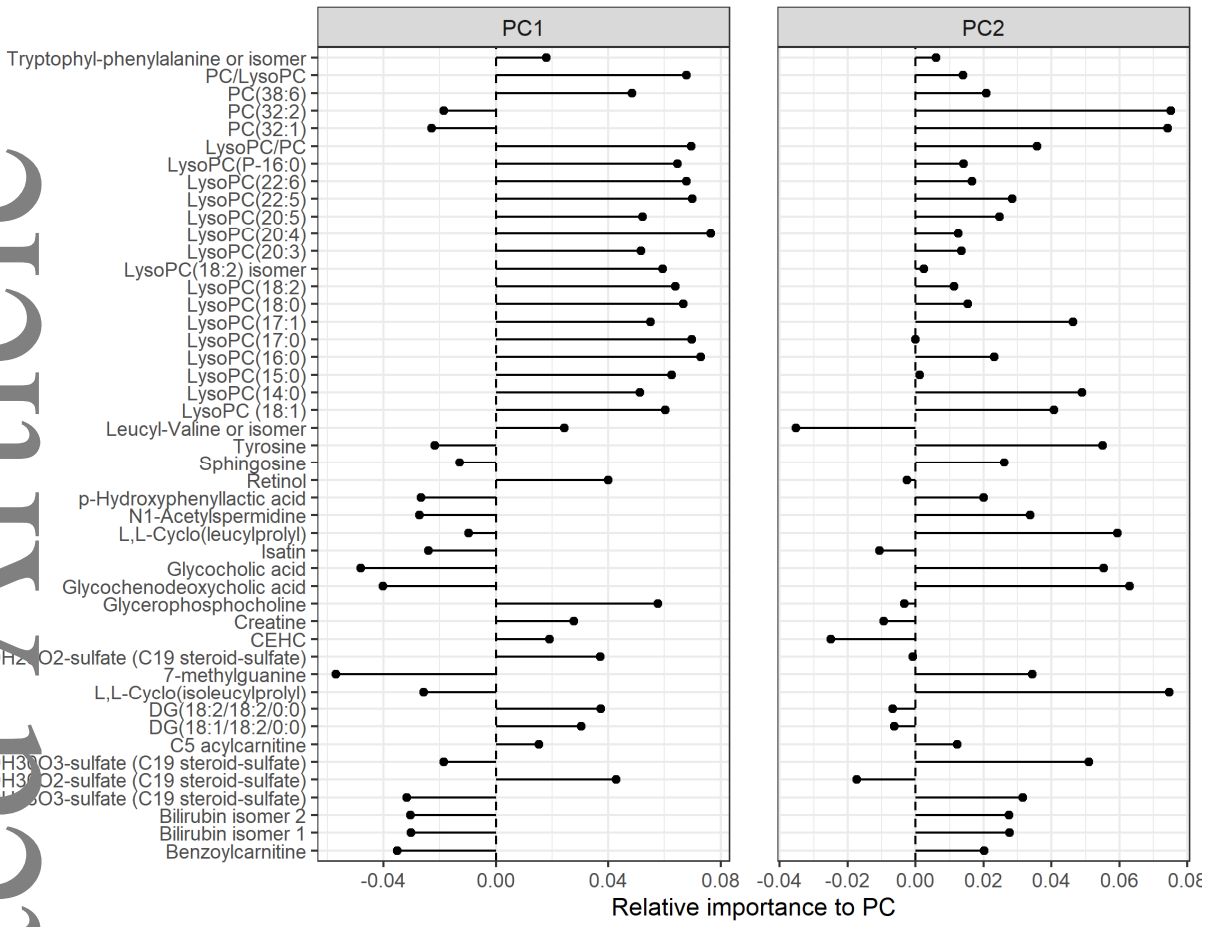


B. Plot of scores on PC1 versus follow-up time

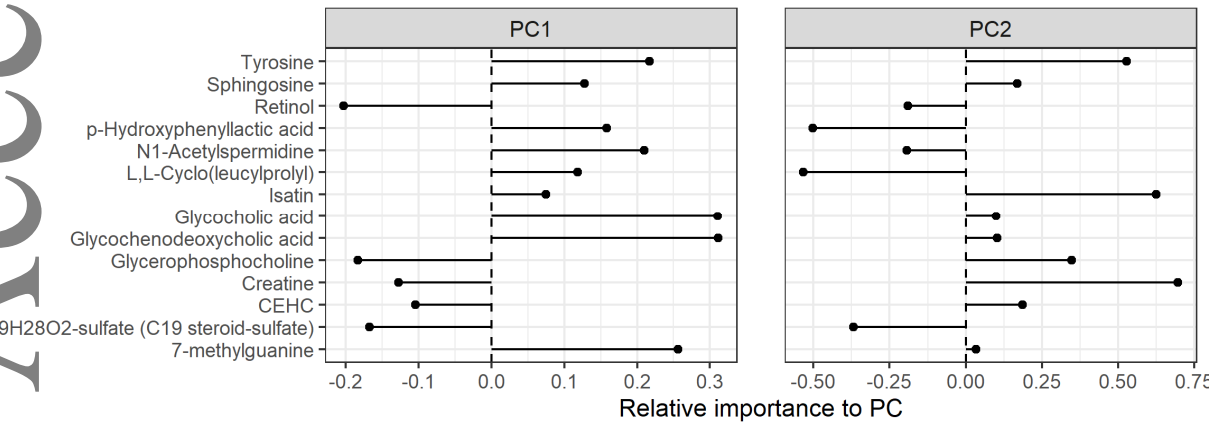


C. Relative contributions of identified metabolites to PC1 and PC2.

(1) 46 metabolites identified at Levels 1-3



(2) 14 metabolites identified at Level 1



THE DIFFERENCE OF BREAKTHROUGH DISCOVERIES ON YOUR TERMS

WITH AN APPROACHABLE, AFFORDABLE, AUTOMATED 4-WAY CELL SORTING SOLUTION IN YOUR LAB. Cell sorting may be complex but it doesn't need to feel complicated or out of reach. With intuitive software that requires minimal training, the BD FACSMelody™ Cell Sorter enables deep scientific insights with reliable results, cost savings and workflow efficiencies. Discover how better instrumentation can free up your time so you can focus your expertise where it matters most. **Discover the new BD.**

Learn how you can advance your research >

BD